



HAL
open science

Interpretability of statistical approaches in speech and language neuroscience

Sophie Bouton, Valerian Chambon, Narly Golestani, Elia Formisano,
Timothée Proix, Anne-Lise Giraud

► **To cite this version:**

Sophie Bouton, Valerian Chambon, Narly Golestani, Elia Formisano, Timothée Proix, et al.. Interpretability of statistical approaches in speech and language neuroscience. 2023. hal-04284936

HAL Id: hal-04284936

<https://hal.science/hal-04284936v1>

Preprint submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 *Interpretability of statistical approaches in speech and language*
2 *neuroscience*

3 Sophie Bouton^{1,2}, Valérian Chambon³, Narly Golestani^{4,5,6}, Elia Formisano^{7,8,9},
4 Timothée Proix¹⁰, Anne-Lise Giraud^{1,10}

5 1. Institut Pasteur, Université Paris Cité, Inserm, Institut de l'Audition, F-75012 Paris,
6 France

7 2. Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Études
8 Cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University, Paris, France

9 3. Institut Jean Nicod, CNRS/École Normale Supérieure UMR 8129, PSL University,
10 75005 Paris, France

11 4. Section of Psychology, University of Geneva – Campus Biotech, 9 chemin des
12 Mines, 1202 Geneva, Switzerland

13 5. Brain and Language Lab, Cognitive Science Hub, University of Vienna, Austria

14 6. Department of Behavioral and Cognitive Biology, Faculty of Life Sciences,
15 University of Vienna, Austria

16 7. Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience,
17 Maastricht University, Maastricht, the Netherlands

18 8. Maastricht Brain Imaging Center, Maastricht, the Netherlands

19 9. Maastricht Centre for Systems Biology, Maastricht University, Maastricht, the
20 Netherlands

21 10. Department of Neuroscience, University of Geneva – Campus Biotech, 9 chemin
22 des Mines, 1202 Geneva, Switzerland

23

24 Timothée Proix and Anne-Lise Giraud contributed equally to this work.

25 The corresponding authors are:

26 **Sophie Bouton**

27 E-Mail address: sophie.bouton@cnr.fr

28 ORCID: <https://orcid.org/0000-0001-5496-4583>

29 **Anne-Lise Giraud**

30 E-Mail address: anne-lise.giraud@unige.ch

31 ORCID: <https://orcid.org/0000-0002-1261-3555>

32 Keywords. data analysis, interpretation, hypothesis-based, data-driven, speech.

33 Abstract

34 The classical view of the speech and language neural system is that of a hierarchy of
35 interdependent modules, enabling the progressive transformation of a continuous
36 acoustic stream into an articulated series of concepts. This modular and hierarchical
37 view follows from the combination of lesion studies (double dissociations) and
38 hypothesis-based factorial designs in which only a few sensory or cognitive factors are
39 varied at a time. In the last ten years, however, data-driven explorations of large
40 neuroimaging datasets have allowed for a more agnostic approach, and led to a
41 whole new view where segregated hierarchically organized modules seem to give way
42 to continuous multidimensional representations, with e.g. a distributed semantic
43 system. While both approaches have brought about significant contributions to
44 speech and language neuroscience, making coherent sense of them represents a
45 substantial challenge. In this review article, we synthesize methodological and
46 experimental findings from the speech and language neuroscience literature,
47 dissecting strength and pitfalls of each approach and suggesting ways in which
48 approaches could be integrated.

49 Main

50 Verification of testable hypotheses serves as the fundamental underpinning of the
51 scientific method. Traditionally, cognitive neuroscience relies on a deductive approach,
52 wherein theoretical models are tested through controlled experiments. However,
53 deductive reasoning alone is often insufficient to make discoveries as the conclusions are
54 implicitly contained in, or restricted to, the hypotheses. In contrast, inductive reasoning
55 involves detecting regularities among observations to infer rules and propose new
56 models. Over the past decade, inductive reasoning relying on the advent of big data and
57 the possibility to train algorithms has increasingly influenced cognitive neuroscience,
58 leading to the discovery of latent structures in the data such as multidimensional and
59 spatially distributed representations. However, *data-driven* analyses alone do not provide
60 a compelling understanding of how these representations relate to cognitive processes,
61 and in particular of how causal they are to behavior. In this article, we outline a
62 methodological taxonomy designed to facilitate the interpretation of results in speech
63 and language neuroscience and foster the integration of the various statistical
64 approaches.

65 **Overcoming Challenges in Integrating Hypothesis-Based and Data-Driven** 66 **Approaches**

67 Traditionally, the hypothesis-based approach in neuroscience consists in relating neural
68 responses to controlled stimuli or behaviors, assuming that only the changes in neural
69 activity associated with one or more pre-specified variables can reliably be interpreted.
70 This approach, based on explicit – mostly univariate – tests of cognitive or psychological
71 models, is often criticized for being constrained to the controlled manipulation of artificial
72 stimuli and conditions with limited generalizability. Yet, it has confirmed the notion that
73 speech-related cognitive operations emerge from modular processing hierarchies where
74 each step has its own computational specificity (Chevillet et al., 2011; Dewitt &
75 Rauschecker, 2012; Formisano et al., 2003; Morillon & Baillet, 2017; Sheng et al., 2018).
76 In the speech and language domain (hereafter referred to as speech), a modular and
77 hierarchically organized neural system enables human speakers to learn the mapping
78 between strings of sounds and meanings, and to combine lexical-level representations in
79 infinite ways to convey new ideas or concepts (Friederici, 2011; Hickok & Poeppel, 2007;
80 Rauschecker & Scott, 2009; Scott & Johnsrude, 2003).

81 In contrast, the data-driven approach makes fewer assumptions about the nature of the
82 system involved, and uses methods that allow the data to be more freely analyzed. Rather
83 than constraining analyses by theoretical knowledge, relationships between features are
84 derived from the data itself. This approach is compatible with more naturalistic stimuli,

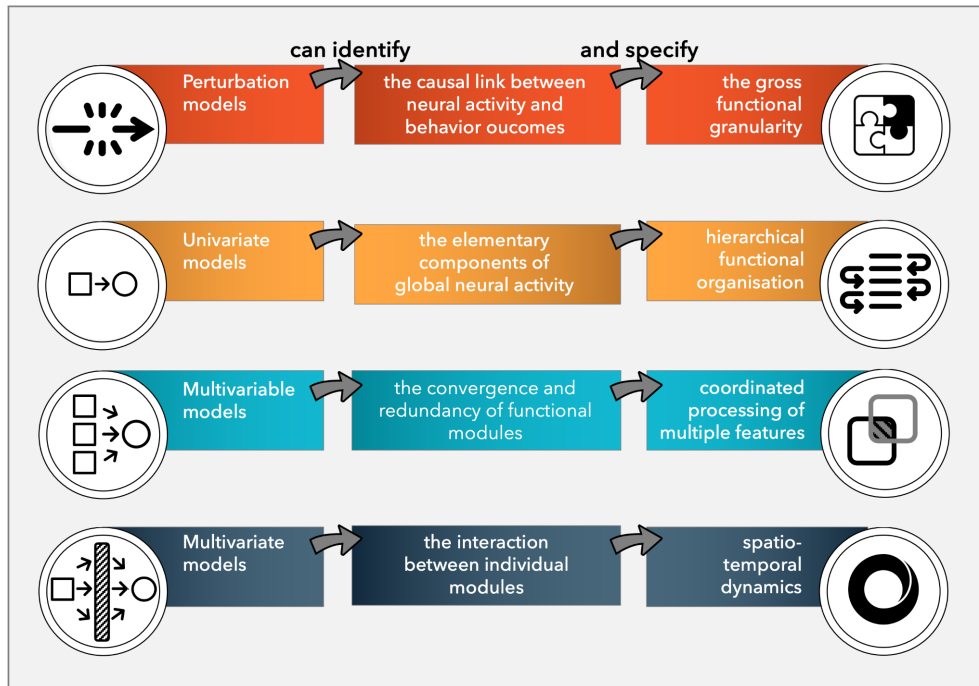
85 and enables exploration of brain responses to language ‘as it is used’ (Hamilton & Huth,
86 2018). However, the use of unsupervised exploration methods and less controlled stimuli
87 considerably obstruct the interpretation of the findings.

88 Reconciling the two approaches seems possible as long as we consider the underlying
89 methods on a continuum, from pure hypothesis-based to fully unsupervised data-driven
90 approaches (Brunton & Beyeler, 2019), rather than as mutually exclusive categories.
91 Likewise, it requires agreeing to provisionally revisit the modular and hierarchical view of
92 speech processing.

93 In what follows, we outline a taxonomy crafted to streamline the integration and
94 interpretation of results yielded by different approaches (**Figure 1**). The taxonomy
95 discerns four statistical approaches: (1) **Perturbation studies** reveal causal relationships
96 between neural activity and behavior outcomes, and specify the gross functional
97 granularity of speech processing. (2) **Univariate models** decompose global neural activity
98 into its elementary components, such as phonemic, syllabic, and semantic features. By
99 paralleling pre-established theories, for instance, those derived from lesion studies, these
100 models probe the hierarchical organization of neural processing. (3) **Multivariable**
101 **models**¹ explore the convergence and redundancy of the aforementioned modules,
102 potentially revealing multiple feature co-processing within the hierarchy. (4) **Multivariate**
103 **models** can be used to examine the interactions between elementary components of the
104 processing hierarchy, thereby identifying the spatio-temporal dynamics that underpins
105 the functional connectivity pattern – an analytical step that specifies how neural elements
106 work together as a system.

107 Drawing on these distinct paradigms, we then highlight the instrumental role of recurrent
108 models. Eschewing a compartmentalized perspective, recurrent models stand out for
109 their capability to identify common elements among the models outlined in our
110 taxonomy. This synthesized-oriented approach aims to harmonize these various
111 approaches, ultimately resulting in a unified model for speech processing.

¹ There is often confusion between the terms multivariable and multivariate (Grant et al., 2019; Hidalgo & Goodman, 2013). Multivariate models refer to the relationship between several dependent variables and one or several independent variables, while multivariable models focus on several independent variables (i.e., multiple features of a stimulus) being regressed onto a single dependent variable (i.e., the neural activity). Thus, multivariable and multivariate models can be used for different purposes.



112

113 **Figure 1.** A taxonomy of the methods used in speech and language neuroscience to integrate
 114 interpretations derived from various statistical approaches. This interpretative guide outlines the
 115 potential benefits of using (1) perturbation models to identify causal links between neural activity and
 116 behavior and to specify gross functional granularity; (2) univariate models to decompose global neural
 117 activity into elementary components and to specify the hierarchical organization; (3) multivariable
 118 models to track the convergence and redundancy of functional modules and to specify coordinated
 119 processing of multiple features; and finally (4) multivariate models to probe interactions between
 120 individual modules and to characterize the underlying spatio-temporal dynamics. Situating each study
 121 within this taxonomy should assist in limiting conclusions to the inferential possibilities of the method
 122 used.

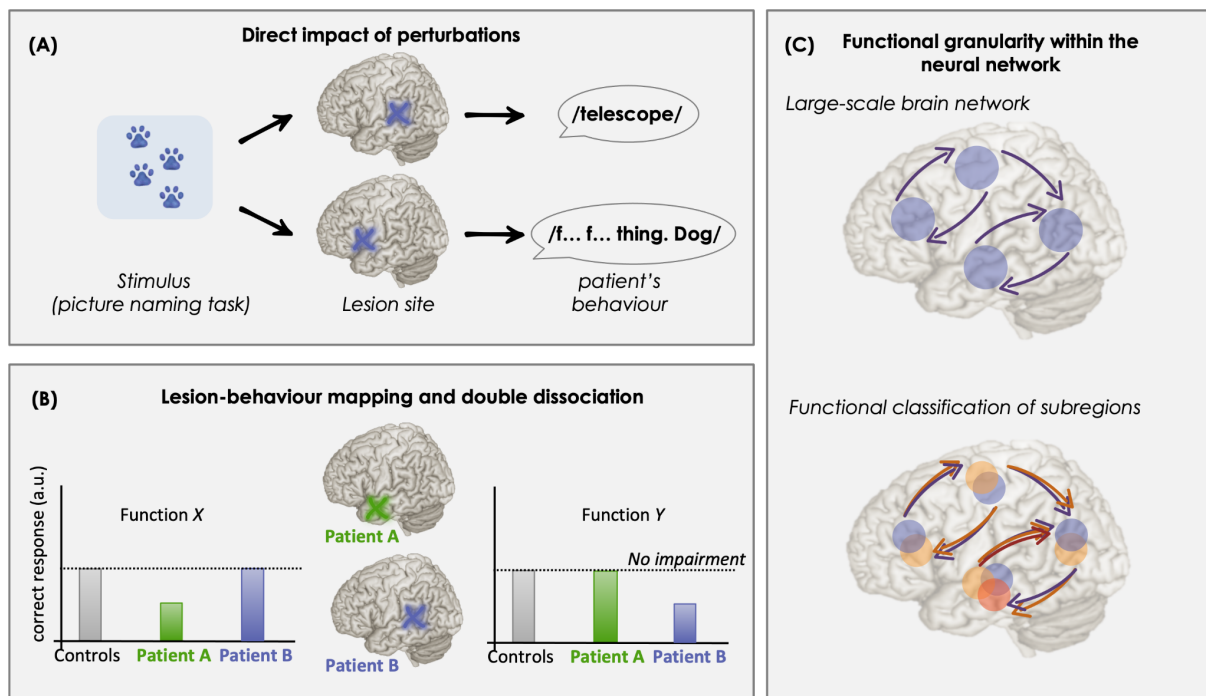
123 **1. Perturbation models to address gross functional granularity**

124 Exploring how brain lesions impact behavior stands as one of the most well-established
 125 and influential methodologies in the field of speech neuroscience. Dating back to the 19th
 126 century, case studies involving patients with focal brain damage demonstrated that
 127 specific language processes are contingent upon distinct brain regions (Broca, 1861;
 128 Wernicke, 1874) (Figure 2A). Over time, lesion-symptom mapping studies, as well as
 129 intervention and perturbation studies, further refined our knowledge of the functional
 130 neuroanatomy of language processes (Dronkers et al., 2017; Penfield & Roberts, 1959)
 131 (Figure 2B). Today, these approaches enable researchers to derive robust observations
 132 on the underlying causal relationships between neural mechanisms and their associated
 133 functions in speech processing (Peters et al., 2017; Weichwald & Jonas, 2021). Crucially,
 134 studies where perturbations are deliberately induced to provoke neural changes are key
 135 to addressing the therapeutic value of an intervention.

136 Perturbation studies can take advantage of chronic or acute disorders such as
137 neurodegenerative atrophy (Chapman et al., 2023; Rogalski et al., 2011) or ischemic
138 stroke (Bouton et al., 2018), alongside clinical interventions such as surgical lesions
139 (Fridriksson et al., 2018; Halai et al., 2018; Hamilton et al., 2021), electrophysiological
140 stimulation (Cardenas et al., 2020; Devlin & Watkins, 2007; Marchesotti et al., 2020; Silva
141 et al., 2022), and cooling (Banerjee et al., 2021; Long et al., 2016). These methods,
142 whether relying on permanent damage or temporary disruptions of neural activity,
143 provide a direct estimate of a specific brain region's role in speech behavior (Fridriksson
144 et al., 2018; Sperber, 2020; Vaidya et al., 2019). Importantly, the clear interpretability of
145 their results is invaluable, especially given the complexities associated with neural
146 processing in speech.

147 In particular, intervention studies have played a pivotal role in elucidating the functional
148 granularity of the speech hierarchy, reconciling inconsistencies arising from correlation
149 studies alone. A noteworthy example pertains to the ongoing debate regarding the
150 functional segregation of lexico-semantic and syntactic processing across different brain
151 modules. While early studies suggest distinct functional modules for syntax and semantics
152 (Hickok & Poeppel, 2007; Matchin & Hickok, 2020), more recent ones highlight the
153 concurrent involvement of a shared brain regions set (Fedorenko et al., 2020). Lesion-
154 behavior mapping has served to reconcile these seemingly incongruous findings by
155 revealing region-specific functional effects (Figure 2C): disorders of syntactic and
156 semantic comprehension manifest subsequent to damages within distinct subparts of the
157 posterior middle temporal gyrus (pMTG) (Matchin et al., 2022).

158 At each functional granularity level, perturbations, whether spontaneous or provoked,
159 allow establishing a causal link between neural processing and behavioral outcomes.
160 While this approach yields clear-cut results, it is, however, hampered by the difficulty of
161 finding pure double dissociation cases that generalize to any new behavioral task,
162 especially for modules higher up the speech hierarchy (Van Orden et al., 2001). Such
163 constraints underscore the potential merit of alternative, data-driven approaches, which
164 do not necessarily hypothesize a strict modular and sequential organization for speech
165 processing but are poised to capture the intricacies of complex neural interactions more
166 effectively.



167

168 **Figure 2.** Leveraging perturbation studies for refined causal granularity of speech and language
 169 processing. (A) Perturbation studies directly probe the impact of local neural activity on a given
 170 behavior. e.g. damage to either the posterior superior temporal gyrus (pSTG) or the inferior frontal
 171 gyrus (IFG) can impair picture naming in different ways: a lesion in the pSTG can lead to fluent but
 172 nonsensical speech, where patients might say a word unrelated to the picture or even create non-
 173 existent words (top); a lesion in the pSTG can lead to non-fluent speech where participants might
 174 struggle to produce the name of the object in the picture, even if they know what it is (bottom). (B) A
 175 crossover double dissociation strengthens the level of evidence, by demonstrating that if damage to
 176 region A causes loss of function X but not Y, and damage to B causes loss of function in Y but not X, X
 177 and Y functions are underpinned by distinct brain modules. (C) Perturbation studies can provide fine-
 178 grained insights into functional modularity within networks of interconnected regions, e.g. by
 179 identifying and describing causal factors operating at different spatial scales within the neural network.

180 2. Univariate models to probe functional specialization of hierarchical modules

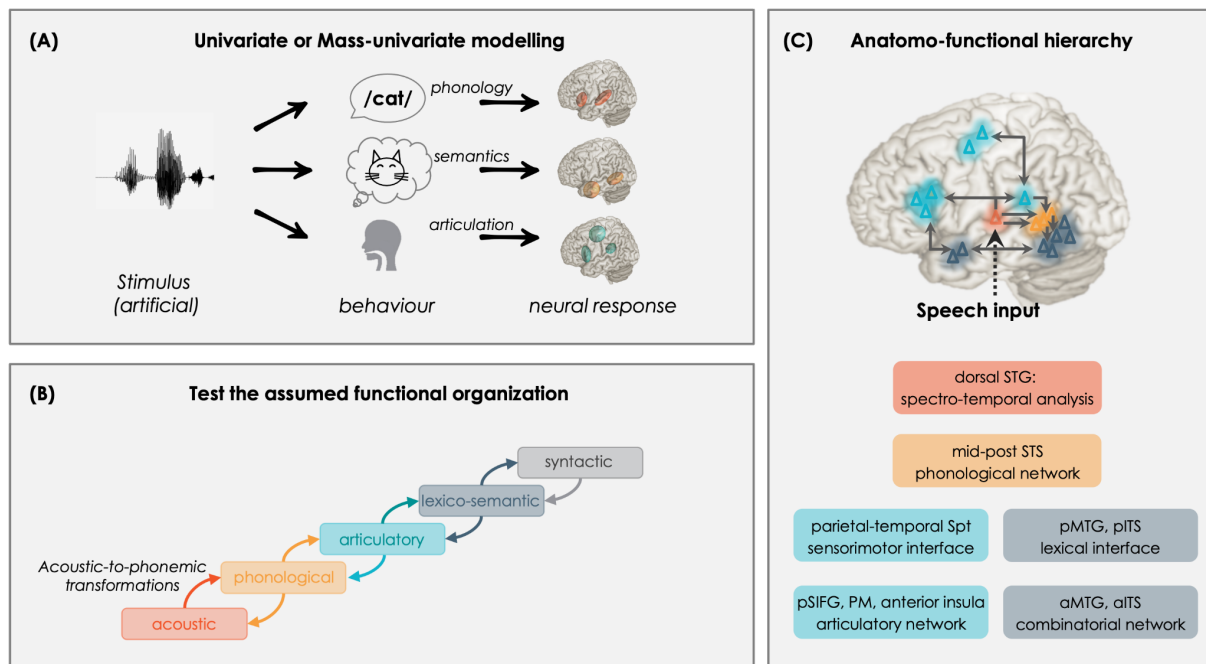
181 A first alternative for more systematic testing of speech functional modules is to relate
 182 changes in neural response to variations in a controlled stimulus, assuming that any
 183 change in the dependent variable (the neural response) is primarily explained by a
 184 controlled manipulation of the independent variable (e.g., the stimulus, the task, the
 185 context) (Figure 3A). The univariate approach is well suited to test predefined hypotheses
 186 about the relationship between specific components of cognitive and psychological
 187 models and the underlying functional neuroanatomy (Figure 3B), as well as about the
 188 functional specificity of each stage in a processing hierarchy (Figure 3C).

189 Univariate encoding models have been widely used to characterize auditory systems, and
 190 provided evidence that information processing proceeds sequentially across
 191 hierarchically organized areas with representations being gradually abstracted along the

192 speech processing hierarchy. As a result, multiple gradients (e.g., tonotopy) and
193 processing levels (core, belt and parabelt) have been identified in the human auditory
194 cortex (Chevillet et al., 2011; Formisano et al., 2003; Talavage et al., 2000). Notably,
195 representations of increasingly complex speech features are organized along an antero-
196 lateral gradient in the temporal cortex, ranging from simple tones and noise bursts to
197 pseudowords and words (Binder, 2000; Dewitt & Rauschecker, 2012). From there, both a
198 ventral and a dorsal stream emerge, providing inputs to the inferior frontal gyrus and
199 reflecting different methods of processing and combining speech units (Friederici, 2011;
200 Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003).

201 Univariate encoding models have also enabled a detailed characterisation of the time
202 scales involved in the speech processing hierarchy (Brennan et al., 2012; Friederici, 2012;
203 Scott, 2000). A time-constrained univariate analysis provided precise temporal neural
204 mapping of multiscale processing during a phonemic categorization task (Bouton et al.,
205 2018). Neural activity associated with low-level perceptual processing was disentangled
206 from higher-level phonemic identification, and selective activity in temporal regions for
207 low-level acoustic features was found 50ms before higher-order decision activity in the
208 left inferior frontal region. In this specific case, the use of univariate time-resolved models
209 allowed for a stepwise spatio-temporal examination of the different computational stages
210 that constitute the speech processing hierarchy, paving the way for a comprehensive
211 understanding of the process.

212 While the hypothesis-based approach, where only one feature varies at a time, enables
213 straightforward interpretation of the results, univariate outcomes from single experiments
214 are insufficient to capture the multi-stream, distributed nature of language processing.
215 They also fall short in untangling the simultaneous contributions of different speech
216 features to the examined neural activity. To achieve a more comprehensive picture of the
217 multi-layered speech processing system, the use of methods adapted to its intrinsic
218 complexity is necessary. This includes multivariable models, which can probe the
219 relationships between multiple variables simultaneously.



220

221 **Figure 3.** Probing speech and language processing hierarchies using univariate models. (A) Univariate,
 222 or mass-univariate, models establish a stimulus-response function relating neural activity to a stimulus
 223 variable, such as an articulatory feature. This hypothesis-based approach uses a few pre-specified
 224 stimulus features or behavioral variables that can be regressed in or of no interest (regressed out). (B)
 225 Univariate models imply assumptions about the relevant variables and assume functional modularity in
 226 the brain. In this view, each region or network of regions primarily encodes specific, well-specified
 227 variables. Historically, the approach derives from the study of sensory systems but is somewhat less
 228 well adapted for characterizing high-level or endogenous functions. (C) Univariate (mass-univariate)
 229 models reveal a neuronal organization, in which low-level information (i.e., spectro-temporal cues), is
 230 transformed along the cortical hierarchy to incrementally build abstract linguistic structures (Figure
 231 adapted from Hickok & Poeppel, 2007).

232 3. Multivariable models to explore multiple feature co-processing

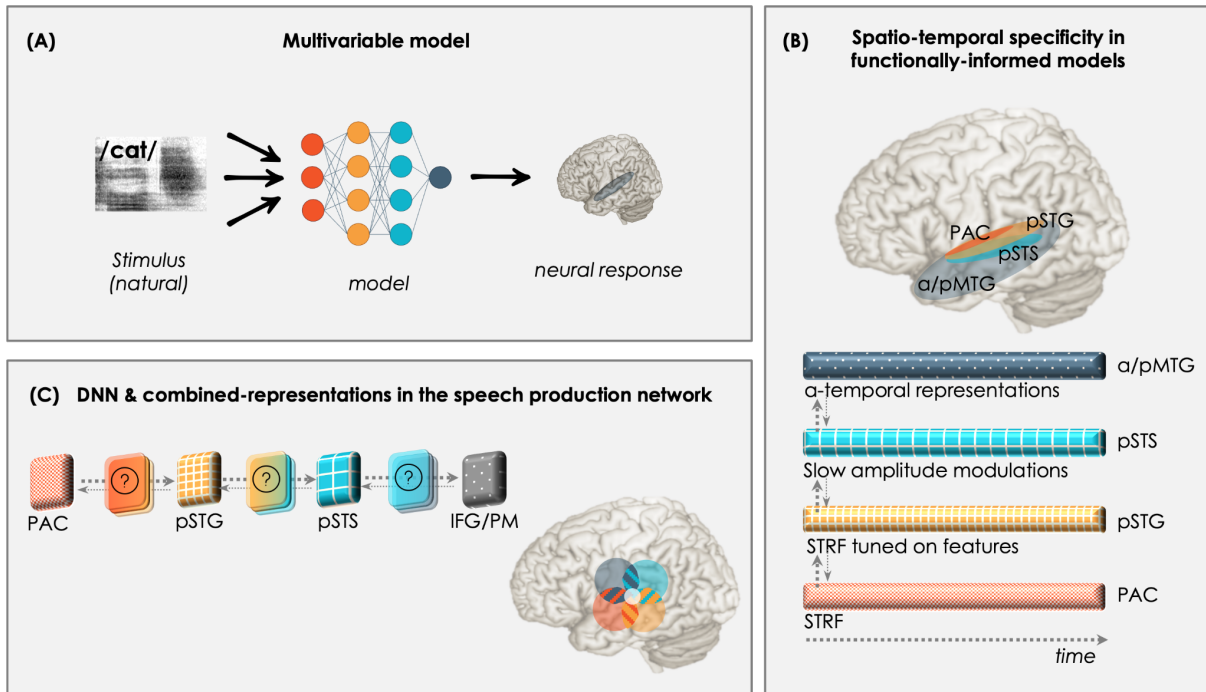
233 Multivariable models have become more prevalent in speech neuroscience research due
 234 to the growing use of naturalistic speech stimuli comprising nested features such as
 235 sentences, words, or syllables, all jointly contributing to a single neural measure.
 236 Multivariable models can disentangle these simultaneous effects by fitting a function that
 237 describes how a set of particular stimulus features contributes to neural responses (Crosse
 238 et al., 2016; Di Liberto et al., 2021) (Figure 4A). Features can be imported from a priori
 239 linguistic assumptions or extracted from data-driven methods such as dimensionality
 240 reduction or large language models fitted to speech data (Caucheteux et al., 2022;
 241 Sainburg et al., 2020; Schrimpf et al., 2021). This approach has confirmed that speech
 242 comprehension relies on hierarchically organized representations, spanning from the
 243 acoustic in primary auditory areas, to phonetic, semantic and higher-level (syntactic)
 244 linguistic representations when moving away from Heschl's gyrus (Caucheteux et al.,

245 2021; de Heer et al., 2017) to prefrontal regions via the double dorsal and ventral stream,
246 a result consistent with previous findings from univariate models.

247 In particular, multivariable models further helped uncover the spatio-temporal
248 specificities of neural activity in the processing hierarchy (Hamilton et al., 2021; Hullett et
249 al., 2016; Palmeri et al., 2017; Pasley et al., 2012; Santoro et al., 2014), as they
250 accommodate the notion that there is an optimal spatio-temporal processing scale at
251 each step considered (Gross et al., 2013; Panzeri et al., 2010; Walker et al., 2011). For
252 example, while a spectro-temporal receptive field (STRF) is the best organization for
253 neural activity in the auditory cortex, an STRF further tuned to phonological cues proves
254 to be a more efficient implementation in the auditory association cortex (Santoro et al.,
255 2017; Venezia et al., 2019). The transformation of the best model throughout the
256 hierarchy reflects the progression from fine temporal resolution in low-level acoustic
257 processing regions, to associative regions of the STG and STS where syllabic and word-
258 level processing regions require integration of longer sequences, and higher-level brain
259 regions (e.g. in the anterior temporal cortex) where semantic representations likely
260 become atemporal (Giraud, 2020), coded with sparse spatial coding (Chang et al., 2010;
261 Deneux et al., 2016; Van Wassenhove, 2009; Q. Zhang et al., 2019). These results
262 highlight the importance of taking into account the hierarchical heterogeneity of
263 functional organization, and of using functionally-informed multivariable models (Figure
264 4B). This description level holds promise for neuroengineering applications, such as
265 developing speech decoders that use neural activity at a particular level of speech
266 processing to decipher distinct components corresponding to specific neural activations
267 (Moses et al., 2019; Tang et al., 2023).

268 More importantly, by enabling direct estimation of concomitant neuronal processing,
269 multivariable models allow for a precise specification of the combined representations of
270 stimulus features at several levels of the hierarchy: mixtures of spectral and articulatory
271 features in the superior temporal gyrus, mixtures of articulatory and semantic features in
272 the superior temporal sulcus (Caucheteux & King, 2022; de Heer et al., 2017), etc. Here,
273 the choice of regressed features is critical as it determines the nature of the information
274 being integrated. For instance, when the multiple variables to be regressed with neural
275 data come from the fitting of a deep neural network (DNN) to behavior, the variables
276 contain implicit information about the architecture and granularity of the hierarchical
277 organization, i.e., N (inter)connected layers (Figure 4C). The number of layers in a DNN
278 is a design choice reflecting the complexity of the task being modeled rather than any
279 biological plausibility. There is hence a risk of smoothing out or missing anatomical
280 (cytoarchitectonic) separations between functionally distinct brain regions when choosing
281 the number of layers, thus creating spurious intermediate stages of processing and

282 representation. Although DNNs draw inspiration from biological neural networks, they
 283 are not direct replicas of biological systems, but rather extreme simplifications (Saxe et
 284 al., 2021). Fitting human behavior with DNNs may thus imply adding extra layers that
 285 have no equivalent in the brain. When seeking to map DNN layers to actual brain activity,
 286 it is therefore necessary to take into account the actual functional granularity of the
 287 cortical hierarchy.



288

289 **Figure 4.** Neural processing involves concomitant neural processing and combined representations.
 290 (A) Multivariable models use multiple independent variables (stimulus features) to predict a single
 291 dependent variable, the neural response. (B) The multivariable model approach implies a priori
 292 assumptions about the spatial and temporal implementation of neural representations. These
 293 assumptions can be functionally informed by the specific computations performed at each level of the
 294 hierarchy. The versatility of these models allows them to address higher-order cognitive operations, so
 295 long as the inherent spatio-temporal characteristics of the associated encoding models are considered.
 296 (C) Mapping deep neural networks (DNNs) to neural data assumes concomitant processing of multiple
 297 features, revealing combined representations. The depth and complexity of DNNs are pivotal in
 298 representing linguistic complexity and abstraction. When inferring about brain-based language
 299 processing, the chosen architecture and layer count of the DNNs should be carefully considered.

300 4. Multivariate models to probe cross-module spatio-temporal dynamics

301 While multivariable models enable the simultaneous analysis of different speech (voice
 302 vs. speech content) and/or linguistic features (phonetic, phonological, lexical etc; see
 303 analysis-by-synthesis, (Poeppel & Monahan, 2011; Pulvermüller & Fadiga, 2010),
 304 identifying multi-network representations of speech requires consideration of processes
 305 distributed across brain regions and neuronal populations (Chalas et al., 2022; Correia et
 306 al., 2015; Lee et al., 2012). These patterns of distributed correlations cannot be revealed

307 with univariate or multivariable models (Ufer & Blank, 2023). In contrast, multivariate
308 models, which relate one or several stimulus features to the activity of multiple
309 simultaneous neural responses (Figure 5A), offer evidence that distributed patterns of
310 neural activity contain discriminative information (Bishop & Nasrabadi, 2006; King et al.,
311 2020).

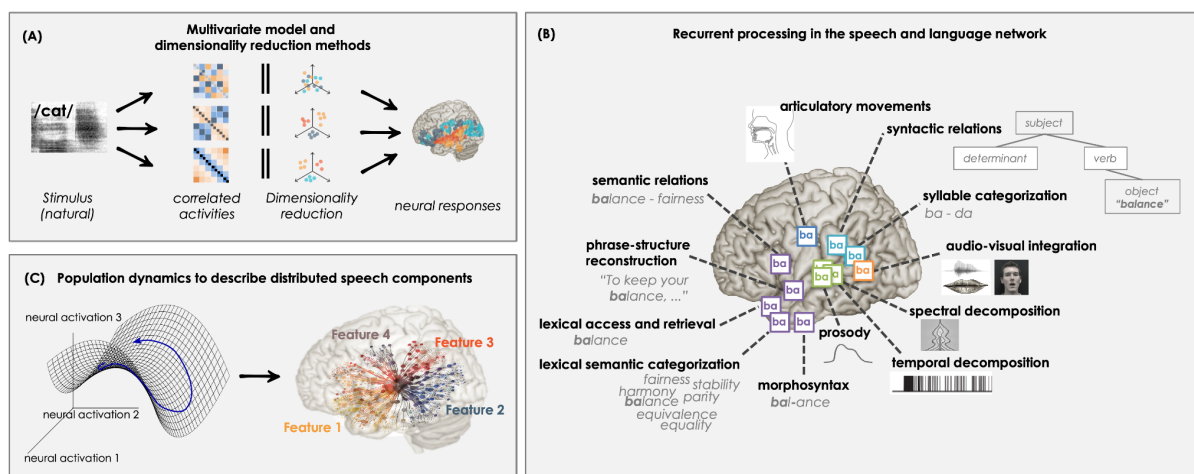
312 Using multivariate models to show that (spatial) patterns of neural activity contain
313 discriminative information proceeds from a different scientific approach than using them
314 in neuroengineering to decode neuronal activity (Hebart & Baker, 2018; Weichwald et al.,
315 2015). Decoding multivariate models aim to reconstruct stimuli from neural activity (K.
316 Friston et al., 2008; Holdgraf et al., 2017; S. Martin et al., 2014) to help palliate deficient
317 functions in patients with, for example, peripheral motor or cortical language diseases
318 (Anumanchipalli et al., 2019; Moses et al., 2021; Willett et al., 2021, 2023). The use of
319 *decoding* multivariate models in cognitive neuroscience requires a cautious approach to
320 prevent misinterpretation of results, as the fitted decoding weights are not directly
321 interpretable as brain activations, but depend on both the signal of interest and the
322 nature of the noise (e.g. neural processing redundancy) present in the signal (Bouton et
323 al., 2018; Haufe et al., 2014; Hebart & Baker, 2018; Kia et al., 2017).

324 Specific methods have been developed to overcome this difficulty. In the linear case, a
325 valuable approach involves inverting fitted multivariate weights to estimate the encoding
326 weights, enabling a direct interpretation of the resulting neural representation of the
327 signal (Haufe et al., 2014). Applied to brain activity during natural speech listening, this
328 method reveals that neural tracking of the speaker's voice pitch is concurrently influenced
329 by acoustic and linguistic features (Kegler et al., 2022). This finding highlights the intricate
330 interactions occurring during the elementary processing steps of speech perception.
331 Another effective solution for interpreting multivariate models is to define regions of
332 interest to spatially constrain the analysis (Çelik et al., 2019). Reliable information on the
333 interactions between levels of elementary processing can be recovered as long as
334 multivariate neural responses are combined at a common spatial resolution (Weichwald
335 et al., 2015). For example, by examining the multivariate fitting performance using all
336 voxels within each brain region of interest, low-level acoustic feature representations were
337 shown to vary depending on whether the subject is engaged in a linguistic or a
338 paralinguistic task (Rutten et al., 2019). This crucial finding aligns with interactive accounts
339 of speech processing and opens up opportunities to further comprehend its dynamics.

340 Alternative approaches aim to improve the interpretability of multivariate analysis in the
341 temporal domain. One such approach seeks to identify the distribution of available
342 information at each point in time. This method can reveal neural reuse, wherein the same
343 brain region's activity is involved sequentially for different purposes. In speech and

344 language processing, such neural reuse might play a crucial role in accumulating sensory
 345 evidence, for example during a dialogue until a response is triggered (Anderson, 2016;
 346 Skipper, 2014). While the use of multivariate methods is critical to ensure that neural reuse
 347 is not missed, recurrent activity is rarely causal to behavior but rather reflects downstream
 348 associations. For instance, spectro-temporal representations can be first activated during
 349 automatic acoustic processing, then used for discriminating one phoneme from another,
 350 and reused to discriminate phonological neighbors (e.g., /bear/-/pear/) (Figure 5B).
 351 Likewise, using multivariate models, letter reading was associated with a distributed
 352 network of successive and overlapping neural activities, implying sequential processing,
 353 maintaining and broadcasting increasingly rich representations across brain regions
 354 (Gwilliams & King, 2020).

355 Finally, at the very end of the hypothesis-based vs data-driven continuum, distributed
 356 speech components can be uncovered using completely unsupervised methods that rely
 357 on a priori measures of interest, such as neural variance, without any relation to sensory
 358 or cognitive features. Dimensionality reduction methods, such as principal component
 359 analysis, can be applied directly to neural data (Bondanelli et al., 2021; Cunningham &
 360 Yu, 2014). These methods reveal low-dimensional spaces of correlated neural activity,
 361 also known as manifolds, that account for a large part of neural variance during speech
 362 tasks. They might provide new insights about how neural activity, specific to each speech
 363 feature and distributed across several brain regions, is integrated into a common
 364 representation (Gallego et al., 2017; Perdakis et al., 2011; Stephen et al., 2023) (Figure
 365 5C). Nevertheless a significant drawback of these methods remains the difficulty of
 366 interpreting the reduced dimensions in functionally meaningful terms.



367
 368 Figure 5. Speech and language processing involve distributed components that can be described
 369 through population dynamics. (A) Multivariate models rely on multiple independent variables to predict
 370 a set of dependent variables, including all possible relevant features of the stimulus. Feature selection
 371 can be achieved using methods such as correlation analyses or dimensionality reduction. (B) Being

372 non-selective, multivariate models can uncover processes of neural reuse. They can also identify
373 distributed networks of sequentially organized neuronal activities. (C) Unsupervised dimensionality
374 reduction method can help reveal underlying population dynamics (left). Projected features can be
375 used to classify neural activity within the speech network (right).

376 **5. Integrating hypothesis-based and data-driven methods**

377 Used in isolation, the analytic methods presented so far enable the identification of
378 distinct speech encoding modes and the characterization of their spatio-temporal
379 distributions. However, these approaches can lead to contradictory conclusions about the
380 functional architecture of speech processing. By combining these models, their individual
381 advantages can be leveraged to get a more unified understanding of speech processing
382 in the brain. In what follows, we highlight the complementary nature of these two
383 approaches, but also how their respective results can inform each other to reveal new
384 principles of neural organization, and finally how these approaches can be integrated
385 together to provide a unified and detailed understanding of speech processing as it
386 unfolds in natural situations.

387 **Complementarity of statistical approaches**

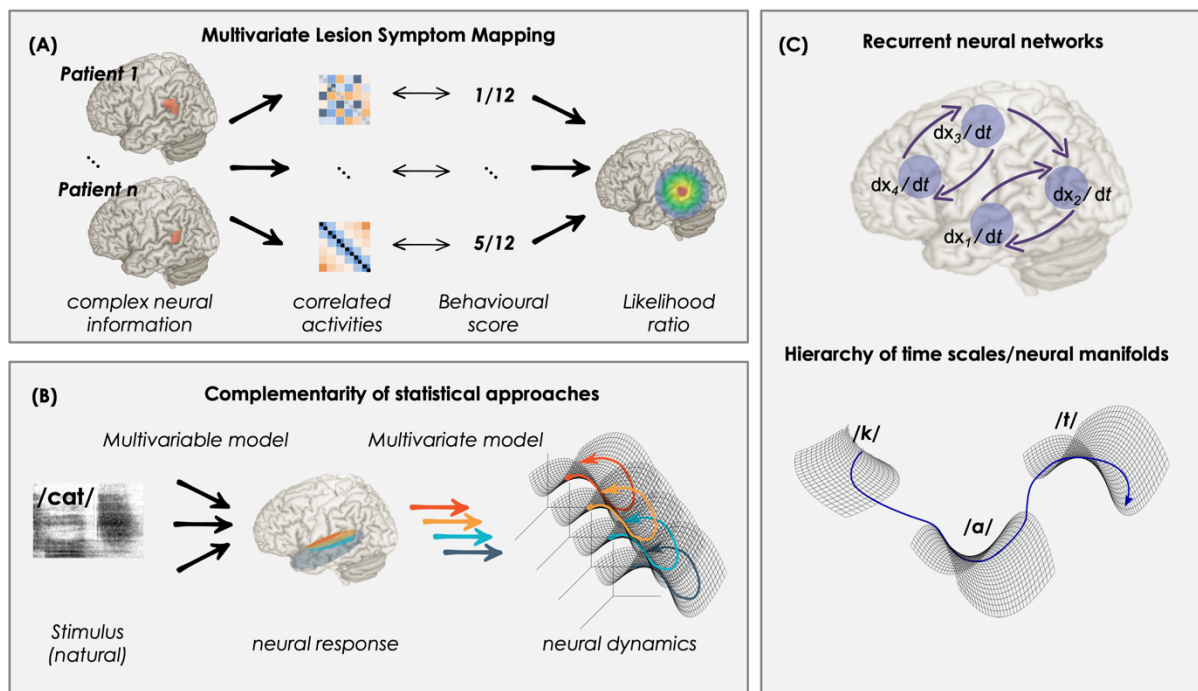
388 Understanding the functional granularity of the speech processing network poses
389 significant challenges when relying on univariate causal models. Complex interactions
390 among interconnected brain regions complicate these efforts (Dronkers et al., 2017;
391 Rahimpour et al., 2019). Indeed, similar speech-deficit symptoms can stem from non-
392 overlapping lesions in different patients (Na et al., 2022), as lesions can not only disrupt
393 local gray-matter regions but also communication between network nodes by damaging
394 white-matter tracts (Geschwind, 1965). Multivariate methods applied to lesion-behavior
395 mapping help overcome these difficulties by taking into account the complex functional
396 interactions between brain regions (Boes et al., 2015; Ivanova et al., 2021; Sperber, 2020;
397 Y. Zhang et al., 2014) (Figure 6A). Lesion network mapping is particularly useful in
398 understanding overall intervention's impact on the neural system by aligning brain lesions
399 with multivariate functional networks (Boes et al., 2015; Fox, 2018; Siddiqi et al., 2022).
400 Crucially, simultaneously introducing lesions and structural connectivity matrices as
401 multivariable regressors can show how functional and structural networks predict
402 behavioral scores. This method has been pivotal in identifying causal network connections
403 and hubs, such as the temporo-parietal junction, a key region for speech fluency, auditory
404 comprehension, speech repetition and oral naming (Yourganov et al., 2016).

405 **Leveraging Hypothesis-Based and Data-Driven Approaches**

406 Interpreting the functional relevance of data-driven reduced dimensions benefits greatly
407 from hypothesis-based methods. This process entails an initial step of pinpointing neural
408 activity associated with specific speech levels, followed by dimensionality reduction to
409 unveil the core spatio-temporal patterns supporting these dynamics (Bouchard et al.,
410 2013; Keshishian et al., 2023; Orepic et al., 2023). Recently this method was applied to
411 demonstrate that the anterior STG simultaneously encodes phonetic and semantic
412 features during natural speech perception, positioning the anterior STG as a promising
413 hub for integrating these two levels of the speech processing hierarchy (Orepic et al.,
414 2023). The synergy of hypothesis-based and data-driven approaches also becomes
415 evident when hypothesis-based findings from controlled experiments are validated using
416 data-driven methods on new datasets collected under varying experimental conditions
417 or during natural conversations (Castellucci et al., 2022; Orepic et al., 2023) (Figure 6B).
418 Relying primarily on hypothesis-based methods, Castellucci et al. identified a speech
419 planning network in controlled experiments. Subsequently, they applied unsupervised
420 dimensionality reduction tools to recover this network in natural conversations,
421 demonstrating its consistency across different conditions.

422 **Integration via recurrent models**

423 To achieve a unified representation of the speech processing system, it is essential to
424 integrate the levels identified (Lupyan & Clark, 2015). This integration must transcend the
425 mere amalgam of speech features and brain regions through causal, multivariable, and
426 multivariate models. Instead, it calls for consideration of how recurrent processes facilitate
427 integration of speech signals over time (Gwilliams et al., 2022; Pulvermüller, 2018; Yi et
428 al., 2019) . This integration requires turning to the concepts of recurrent and dynamical
429 models, where past neural activity shapes future neural dynamics (K. J. Friston et al., 2003;
430 Truccolo et al., 2004; Vyas et al., 2020). Recurrent models make it possible to examine
431 multiple variables at once, and how different brain regions and speech features influence
432 neural activity, while preserving causality, interpretability, and testability (Figure 6C).
433 When applied to auditory cortex neural activity, recurrent models reveal how different
434 auditory contents are represented in the brain by exposing hidden low-dimensional,
435 dynamical structures (Bondanelli et al., 2021). Dynamic recurrent models have the
436 intriguing feature of allowing the definition of distinct time constants for various sub-
437 processes, thus forming a hierarchy of time scales that effectively integrates the distinct
438 stages of the speech processing hierarchy (A. E. Martin, 2020; Perdikis et al., 2011; Pillai
439 & Jirsa, 2017). More generally, this framework has the potential to characterize neural
440 activity during tasks beyond speech, such as word writing or similar dynamic tasks
441 (Perdikis et al., 2011).



442

443 **Figure 6.** Speech processing involves integrating representations through recurrent activations. (A)
 444 Multivariable lesion symptom mapping, which consists of introducing multiple lesioned voxels or
 445 connectivity measurements as independent variables, allows for the consideration of the effect of the
 446 lesioned speech network on behavioral deficits. Multivariable lesion symptom mapping does not
 447 require patients to be grouped by either lesion site or behavioral cutoff, but instead makes use of
 448 continuous behavioral and lesion information. (B) Two-steps analysis helps in interpreting distributed
 449 neural patterns obtained after applying dimensionality reduction: first, an encoding multivariable
 450 model is applied to identify which neural activity relates to which speech features across the whole
 451 recording, and secondly, the resulting dimensionality is reduced to ease interpretation. (C) Recurrent
 452 models allow for setting dynamics at different time scales that integrate speech features across levels
 453 of the hierarchy.

454 **Conclusion**

455 We propose that the speech and language function in the human brain can be fully
 456 characterized by five cardinal approaches. These complementary approaches illustrate
 457 the strengths and limitations of hypothesis-based and data-driven methods. While
 458 perturbation models demonstrate causal links between neural processes and behavior,
 459 providing a clear understanding of the speech processing modularity, univariate models
 460 allow global neural activity to be broken down into elementary components or modules,
 461 improving our understanding of the hierarchical organization of neural processing.
 462 Multivariable models complement this approach by tracking multiple feature
 463 convergences and identifying edges with shared representations within the hierarchy.
 464 Multivariate models can reveal spatially and temporally distributed components and their

465 dynamics. Finally, recurrent models give an account of how different speech features are
466 integrated over time within the speech hierarchy.

467

468 This article stresses the need to integrate the exploratory power of data-driven methods
469 with the hypothetico-deductive approach. Indeed, data-driven research is constantly
470 navigating between the extraordinary possibilities it offers for new discoveries and the
471 pitfalls of random-walk science. Attention to past research and theoretical concerns can
472 help prevent this random walk from getting lost in a vast, ultimately uninteresting space.
473 This tension between deductive and inductive approaches is far from new in science. Two
474 decades ago, science-fiction writer Ted Chiang wrote a premonitory article in *Nature* that
475 purported to celebrate the 25th anniversary of the last human scientific publication
476 (Chiang, 2000). In this provocative futuristic short-story, science is exclusively conducted
477 by genetically improved humans, called "metahumans", and all that remains for human
478 scientists is the interpretation of meta-human data. It is tempting to think of metahuman
479 science as the outcome of data-/AI-driven methods, whose inner functioning lies
480 sometimes beyond the reach of our understanding. But data are rarely, if ever, self-
481 explanatory. They make sense to the extent that they validate or falsify theoretically
482 informed predictions. "Human researchers", Chiang says somewhat presciently, "may
483 indeed discern applications overlooked by metahumans, whose advantages tend to make
484 them unaware of *our* concerns".

485 References

- 486 Anderson, M. L. (2016). Précis of *After Phrenology: Neural Reuse and the Interactive Brain*. *Behavioral and Brain*
487 *Sciences*, 39, e120. <https://doi.org/10.1017/S0140525X15000631>
- 488 Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken
489 sentences. *Nature*, 568(7753), 493–498. <https://doi.org/10.1038/s41586-019-1119-1>
- 490 Banerjee, A., Egger, R., & Long, M. A. (2021). Using focal cooling to link neural dynamics and behavior. *Neuron*,
491 109(16), 2508–2518. <https://doi.org/10.1016/j.neuron.2021.05.029>
- 492 Binder, J. R. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*,
493 10(5), 512–528. <https://doi.org/10.1093/cercor/10.5.512>
- 494 Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, Issue 4). Springer.
- 495 Boes, A. D., Prasad, S., Liu, H., Liu, Q., Pascual-Leone, A., Caviness, V. S., & Fox, M. D. (2015). Network
496 localization of neurological symptoms from focal brain lesions. *Brain*, 138(10), 3061–3075.
497 <https://doi.org/10.1093/brain/awv228>
- 498 Bondanelli, G., Deneux, T., Bathellier, B., & Ostojic, S. (2021). Network dynamics underlying OFF responses in
499 the auditory cortex. *eLife*, 10, e53151. <https://doi.org/10.7554/eLife.53151>
- 500 Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human
501 sensorimotor cortex for speech articulation. *Nature*, 495(7441), 327–332.
502 <https://doi.org/10.1038/nature11911>
- 503 Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A. G., Seeck, M., Karkar, S., Ville, D. V. D., & Giraud, A.-L.
504 (2018). Focal versus distributed temporal cortex activity for speech sound category assignment. *PNAS*,
505 115(6), E1299–E1308. <https://doi.org/10.1101/133272>
- 506 Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pykkänen, L. (2012). Syntactic structure building in
507 the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2), 163–173.
508 <https://doi.org/10.1016/j.bandl.2010.04.002>
- 509 Broca, P. (1861). Nouvelle observation d'aphémie produite par une lésion de la troisième circonvolution frontale.
510 *Bulletins de La Société D'anatomie*, 6(2), 398–407.
- 511 Brunton, B. W., & Beyeler, M. (2019). Data-driven models in human neuroscience and neuroengineering. *Current*
512 *Opinion in Neurobiology*, 58, 21–29. <https://doi.org/10.1016/j.conb.2019.06.008>

- 513 Cardenas, A. R., Behroozmand, R., Kocsis, Z., Gander, P. E., Nourski, K. V., Kovach, C. K., Ibayashi, K., Pipoly,
514 M., Kawasaki, H., Howard, M. A., & Greenlee, J. D. (2020). *Differential causal involvement of human*
515 *auditory and frontal cortices in vocal motor control* [Preprint]. *Neuroscience*.
516 <https://doi.org/10.1101/2020.06.08.139881>
- 517 Castellucci, G. A., Kovach, C. K., Howard, M. A., Greenlee, J. D. W., & Long, M. A. (2022). A speech planning
518 network for interactive language use. *Nature*. <https://doi.org/10.1038/s41586-021-04270-z>
- 519 Caucheteux, C., Gramfort, A., & King, J.-R. (2021). Model-based analysis of brain activity reveals the hierarchy of
520 language in 305 subjects. *arXiv*, 10. arXiv. <https://doi.org/2110.06078>
- 521 Caucheteux, C., Gramfort, A., & King, J.-R. (2022). Deep language algorithms predict semantic comprehension
522 from brain activity. *Scientific Reports*, 12(1), 16327. <https://doi.org/10.1038/s41598-022-20460-9>
- 523 Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing.
524 *Communications Biology*, 5(1), 134. <https://doi.org/10.1038/s42003-022-03036-1>
- 525 Çelik, E., Dar, S. U. H., Yilmaz, Ö., Keleş, Ü., & Çukur, T. (2019). Spatially informed voxelwise modeling for
526 naturalistic fMRI experiments. *NeuroImage*, 186, 741–757.
527 <https://doi.org/10.1016/j.neuroimage.2018.11.044>
- 528 Chalas, N., Daube, C., Kluger, D. S., Abbasi, O., Nitsch, R., & Gross, J. (2022). Multivariate analysis of speech
529 envelope tracking reveals coupling beyond auditory cortex. *NeuroImage*, 258, 119395.
530 <https://doi.org/10.1016/j.neuroimage.2022.119395>
- 531 Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical
532 Speech Representation in the Human Superior Temporal Gyrus. *Nature Neuroscience*, 13(11), 1428–
533 1432. <https://doi.org/10.1038/nn.2641.Categorical>
- 534 Chapman, C. A., Polyakova, M., Mueller, K., Weise, C., Fassbender, K., Fliessbach, K., Kornhuber, J., Lauer, M.,
535 Anderl-Straub, S., Ludolph, A., Prudlo, J., Staiger, A., Synofzik, M., Wilfang, J., Riedl, L., Diehl-Schmid,
536 J., Otto, M., Danek, A., FTL D Consortium Germany, ... Schroeter, M. L. (2023). Structural correlates of
537 language processing in primary progressive aphasia. *Brain Communications*, 5(2), fcad076.
538 <https://doi.org/10.1093/braincomms/fcad076>
- 539 Chevillet, M., Riesenhuber, M., & Rauschecker, J. P. (2011). Functional Correlates of the Anterolateral
540 Processing Hierarchy in Human Auditory Cortex. *Journal of Neuroscience*, 31(25), 9345–9352.
541 <https://doi.org/10.1523/JNEUROSCI.1448-11.2011>
- 542 Chiang, T. (2000). Catching crumbs from the table. *Nature*, 405, 517.
- 543 Correia, J. M., Jansma, B. M. B., & Bonte, M. (2015). Decoding Articulatory Features from fMRI Responses in
544 Dorsal Speech Regions. *The Journal of Neuroscience*, 35(45), 15015–15025.
545 <https://doi.org/10.1523/JNEUROSCI.0977-15.2015>
- 546 Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function
547 (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in*
548 *Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00604>
- 549 Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature*
550 *Neuroscience*, 17(11), 1500–1509. <https://doi.org/10.1038/nn.3776>
- 551 de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical
552 organization of human speech processing. *Journal of Neuroscience*, 37(27), 6539–6557.
553 <https://doi.org/10.1523/JNEUROSCI.3267-16.2017>
- 554 Deneux, T., Kempf, A., Daret, A., Ponsot, E., & Bathellier, B. (2016). Temporal asymmetries in auditory coding
555 and perception reflect multi-layered nonlinearities. *Nature Communications*, 7(1), 12682.
556 <https://doi.org/10.1038/ncomms12682>
- 557 Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain*, 130(3), 610–622.
558 <https://doi.org/10.1093/brain/awl331>
- 559 Dewitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *PNAS*,
560 109(8), 505–514. <https://doi.org/10.1073/pnas.1113427109>
- 561 Di Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural
562 representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, 227.
563 <https://doi.org/10.1016/j.neuroimage.2020.117586>
- 564 Dronkers, N. F., Ivanova, M. V., & Baldo, J. V. (2017). What Do Language Disorders Reveal about Brain–
565 Language Relationships? From Classic Models to Network Approaches. *Journal of the International*
566 *Neuropsychological Society*, 23(9–10), 741–754. <https://doi.org/10.1017/S1355617717001126>
- 567 Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word
568 meanings throughout the language network. *Cognition*, 203, 104348.
569 <https://doi.org/10.1016/j.cognition.2020.104348>
- 570 Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., & Goebel, R. (2003). Mirror-Symmetric
571 Tonotopic Maps in Human Primary Auditory Cortex. *Neuron*, 40(4), 859–869.
572 [https://doi.org/10.1016/S0896-6273\(03\)00669-X](https://doi.org/10.1016/S0896-6273(03)00669-X)
- 573 Fox, M. D. (2018). Mapping Symptoms to Brain Networks with the Human Connectome. *New England Journal of*
574 *Medicine*, 379(23), 2237–2245. <https://doi.org/10.1056/NEJMra1706158>
- 575 Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., & Bonilha, L.
576 (2018). Anatomy of aphasia revisited. *Brain*, 141(3), 848–862. <https://doi.org/10.1093/brain/awx363>
- 577 Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological*
578 *Reviews*, 91(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>
- 579 Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension.

580 *Trends in Cognitive Sciences*, 16(5), 262–268. <https://doi.org/10.1016/j.tics.2012.04.001>

581 Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., & Ashburner, J. (2008). Bayesian
582 decoding of brain images. *NeuroImage*, 39(1), 181–205.
583 <https://doi.org/10.1016/j.neuroimage.2007.08.013>

584 Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302.
585 [https://doi.org/10.1016/S1053-8119\(03\)00202-7](https://doi.org/10.1016/S1053-8119(03)00202-7)

586 Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement.
587 *Neuron*, 94(5), 978–984. <https://doi.org/10.1016/j.neuron.2017.05.025>

588 Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain: A Journal of Neurology*, 88(2),
589 237–294. <https://doi-org.inshs.bib.cnrs.fr/10.1093/brain/88.2.237>

590 Giraud, A.-L. (2020). Oscillations for all \backslash (\surd) $_$ $_$? A commentary on Meyer, Sun & Martin (2020). *Language,*
591 *Cognition and Neuroscience*, 35(9), 1106–1113. <https://doi.org/10.1080/23273798.2020.1764990>

592 Grant, S. W., Hickey, G. L., & Head, S. J. (2019). Statistical primer: Multivariable regression considerations and
593 pitfalls†. *European Journal of Cardio-Thoracic Surgery*, 55(2), 179–185.
594 <https://doi.org/10.1093/ejcts/ezy403>

595 Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms
596 and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, 11(12), e1001752.
597 <https://doi.org/10.1371/journal.pbio.1001752>

598 Gwilliams, L., & King, J.-R. (2020). Recurrent processes support a cascade of hierarchical decisions. *eLife*, 9, 1–
599 20. <https://doi.org/10.7554/eLife.56603>

600 Gwilliams, L., King, J.-R., Marantz, A., & Poeppel, D. (2022). Neural dynamics of phoneme sequences reveal
601 position-invariant code for content and order. *Nature Communications*, 13(1), Article 1.
602 <https://doi.org/10.1038/s41467-022-34326-1>

603 Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2018). Triangulation of language-cognitive impairments,
604 naming errors and their neural bases post-stroke. *NeuroImage: Clinical*, 17, 465–473.
605 <https://doi.org/10.1016/j.nicl.2017.10.037>

606 Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: Natural stimuli in speech
607 neuroscience. *Language, Cognition and Neuroscience*, 35(5), 573–582.
608 <https://doi.org/10.1080/23273798.2018.1499946>

609 Hamilton, L. S., Oganian, Y., Hall, J., & Chang, E. F. (2021). Parallel and distributed encoding of speech across
610 human auditory cortex. *Cell*, 184(18), 4626–4639.e13. <https://doi.org/10.1016/j.cell.2021.07.019>

611 Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the
612 interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87, 96–110.
613 <https://doi.org/10.1016/j.neuroimage.2013.10.067>

614 Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function.
615 *NeuroImage*, 180, 4–18. <https://doi.org/10.1016/j.neuroimage.2017.08.005>

616 Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*,
617 8(5), 393–402. <https://doi.org/10.1038/nrn2113>

618 Hidalgo, B., & Goodman, M. (2013). Multivariate or Multivariable Regression? *American Journal of Public Health*,
619 103(1), 39–40. <https://doi.org/10.2105/AJPH.2012.300897>

620 Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., Theunissen, F. E., & David, S. V. (2017).
621 Encoding and Decoding Models in Cognitive Electrophysiology. *Frontiers in Systems Neuroscience*, 11,
622 61. <https://doi.org/10.3389/fnsys.2017.00061>

623 Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human Superior Temporal
624 Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. *The Journal of*
625 *Neuroscience*, 36(6), 2014–2026. <https://doi.org/10.1523/JNEUROSCI.1779-15.2016>

626 Ivanova, M. V., Herron, T. J., Dronkers, N. F., & Baldo, J. V. (2021). An empirical comparison of univariate versus
627 multivariate methods for the analysis of brain–behavior mapping. *Human Brain Mapping*, 42(4), 1070–
628 1101. <https://doi.org/10.1002/hbm.25278>

629 Kegler, M., Weissbart, H., & Reichenbach, T. (2022). The neural response at the fundamental frequency of
630 speech is modulated by word-level acoustic and linguistic information. *Frontiers in Neuroscience*, 16,
631 915744. <https://doi-org.inshs.bib.cnrs.fr/10.3389/fnins.2022.915744>

632 Keshishian, M., Akkol, S., Herrero, J., Bickel, S., Mehta, A. D., & Mesgarani, N. (2023). Joint, distributed and
633 hierarchically organized encoding of linguistic features in the human auditory cortex. *Nature Human*
634 *Behaviour*. <https://doi.org/10.1038/s41562-023-01520-0>

635 Kia, S. M., Vega Pons, S., Weisz, N., & Passerini, A. (2017). Interpretability of Multivariate Brain Maps in Linear
636 Brain Decoding: Definition, and Heuristic Quantification in Multivariate Analysis of MEG Time-Locked
637 Effects. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00619>

638 King, J.-R., Gwilliams, L., Holdgraf, C. R., Sassenhagen, J., Barachant, A., Engemann, D., Larson, E., &
639 Gramfort, A. (2020). Encoding and Decoding Neuronal Dynamics: Methodological Framework to
640 Uncover the Algorithms of Cognition. In *The Cognitive Neurosciences, 6th edition* (Poeppel, D., Mangun,
641 G.R., Gazzaniga, M.S., pp. 691–702). The MIT Press.

642 Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical Speech Processing in Broca's
643 Area: An fMRI Study Using Multivariate Pattern-Based Analysis. *The Journal of Neuroscience*, 32(11),
644 3942–3948. <https://doi.org/10.1523/JNEUROSCI.3814-11.2012>

645 Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., Oya, H., Howard, M. A., &
646 Greenlee, J. D. W. (2016). Functional Segregation of Cortical Regions Underlying Speech Timing and

647 Articulation. *Neuron*, 89(6), 1187–1193. <https://doi.org/10.1016/j.neuron.2016.01.032>

648 Lupyán, G., & Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition
649 Interface. *Current Directions in Psychological Science*, 24(4), 279–284.
650 <https://doi.org/10.1177/0963721415570732>

651 Marchesotti, S., Nicolle, J., Merlet, I., Arnal, L. H., Donoghue, J. P., & Giraud, A.-L. (2020). Selective
652 enhancement of low-gamma activity by tACS improves phonemic processing and reading accuracy in
653 dyslexia. *PLOS Biology*, 18(9), e3000833. <https://doi.org/10.1371/journal.pbio.3000833>

654 Martin, A. E. (2020). A Compositional Neural Architecture for Language. *Journal of Cognitive Neuroscience*,
655 32(8), 1407–1427. https://doi.org/10.1162/jocn_a_01552

656 Martin, S., Brunner, P., Holdgraf, C. R., Heinze, H., & Crone, N. E. (2014). Decoding spectrotemporal features of
657 overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, 7, 1–15.
658 <https://doi.org/10.3389/fneng.2014.00014>

659 Matchin, W., Basilakos, A., Ouden, D.-B. den, Stark, B. C., Hickok, G., & Fridriksson, J. (2022). Functional
660 differentiation in the language network revealed by lesion-symptom mapping. *NeuroImage*, 247, 118778.
661 <https://doi.org/10.1016/j.neuroimage.2021.118778>

662 Matchin, W., & Hickok, G. (2020). The Cortical Organization of Syntax. *Cerebral Cortex*, 30(3), 1481–1498.
663 <https://doi.org/10.1093/cercor/bhz180>

664 Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the
665 National Academy of Sciences*, 114(42), E8913–E8921. <https://doi.org/10.1073/pnas.1705373114>

666 Moses, D. A., Leonard, M. K., Makin, J. G., & Chang, E. F. (2019). Real-time decoding of question-and-answer
667 speech dialogue using human cortical activity. *Nature Communications*, 10(1), 3096.
668 <https://doi.org/10.1038/s41467-019-10994-4>

669 Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty,
670 M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for
671 Decoding Speech in a Paralyzed Person with Anarthria. *New England Journal of Medicine*, 385(3), 217–
672 227. <https://doi.org/10.1056/NEJMoa2027540>

673 Na, Y., Jung, J., Tench, C. R., Auer, D. P., & Pyun, S.-B. (2022). Language systems from lesion-symptom
674 mapping in aphasia: A meta-analysis of voxel-based lesion mapping studies. *NeuroImage: Clinical*, 35,
675 103038. <https://doi.org/10.1016/j.nicl.2022.103038>

676 Orepic, P., Truccolo, W., Cash, S. S., Giraud, A.-L., & Proix, T. (2023). Low-dimensional neuronal population
677 dynamics in anterior superior temporal gyrus reactivate phonetic representations during semantic
678 processing [Preprint]. *Neuroscience*. <https://doi.org/10.1101/2023.10.30.564638>

679 Palmeri, T. J., Love, B. C., & Turner, B. M. (2017). Model-based cognitive neuroscience. *Journal of Mathematical
680 Psychology*, 76, 59–64. <https://doi.org/10.1016/j.jmp.2016.10.010>

681 Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2010). Sensory neural codes using multiplexed temporal
682 scales. *Trends in Neurosciences*, 33(3), 111–120. <https://doi.org/10.1016/j.tins.2009.12.001>

683 Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., & Chang, E. F.
684 (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1), e1001251.
685 <https://doi.org/10.1371/journal.pbio.1001251>

686 Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton University Press.

687 Perdikis, D., Huys, R., & Jirsa, V. K. (2011). Time Scale Hierarchies in the Functional Organization of Complex
688 Behaviors. *PLoS Computational Biology*, 7(9), e1002198. <https://doi.org/10.1371/journal.pcbi.1002198>

689 Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: Foundations and learning
690 algorithms*. MIT Press.

691 Pillai, A. S., & Jirsa, V. K. (2017). Symmetry Breaking in Space-Time Hierarchies Shapes Brain Dynamics and
692 Behavior. *Neuron*, 94(5), 1010–1026. <https://doi.org/10.1016/j.neuron.2017.05.013>

693 Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by
694 synthesis. *Language and Cognitive Processes*, 26(7), 935–951.
695 <https://doi.org/10.1080/01690965.2010.493301>

696 Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication.
697 *Progress in Neurobiology*, 160, 1–44. <https://doi.org/10.1016/j.pneurobio.2017.07.001>

698 Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language.
699 *Nature Reviews Neuroscience*, 11(5), 351–360.

700 Rahimpour, S., Haglund, M. M., Friedman, A. H., & Duffau, H. (2019). History of awake mapping and speech and
701 language localization: From modules to networks. *Neurosurgical Focus*, 47(3), E4.
702 <https://doi.org/10.3171/2019.7.FOCUS19347>

703 Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates
704 illuminate human speech processing. *Nature Neuroscience*, 12(6), 718–724.
705 <https://doi.org/10.1038/nn.2331>

706 Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Weintraub, S., & Mesulam, M.-M. (2011). Progression of
707 language decline and cortical atrophy in subtypes of primary progressive aphasia. *Neurology*, 76(21),
708 1804–1810. <https://doi.org/10.1212/WNL.0b013e31821ccd3c>

709 Rutten, S., Santoro, R., Hervais-Adelman, A., Formisano, E., & Golestani, N. (2019). Cortical encoding of speech
710 enhances task-relevant acoustic information. *Nature Human Behaviour*, 3(9), 974–987.
711 <https://doi.org/10.1038/s41562-019-0648-9>

712 Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across
713 diverse animal vocal repertoires. *PLOS Computational Biology*, 16(10), e1008228.

714 <https://doi.org/10.1371/journal.pcbi.1008228>

715 Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of
716 Natural Sounds at Multiple Spectral and Temporal Resolutions in the Human Auditory Cortex. *PLoS*
717 *Computational Biology*, 10(1), e1003412. <https://doi.org/10.1371/journal.pcbi.1003412>

718 Santoro, R., Moerel, M., Martino, F. D., Valente, G., Ugurbil, K., Yacoub, E., & Formisano, E. (2017).
719 Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns.
720 *Proceedings of the National Academy of Sciences*, 114(18), 4799–4804.
721 <https://doi.org/10.1073/pnas.1617622114>

722 Saxe, A., Nelli, S., & Summerfield, C. (2021). If deep learning is the answer, what is the question? *Nature*
723 *Reviews Neuroscience*, 22(1), 55–67. <https://doi.org/10.1038/s41583-020-00395-8>

724 Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko,
725 E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing.
726 *Proceedings of the National Academy of Sciences*, 118(45), e2105646118.
727 <https://doi.org/10.1073/pnas.2105646118>

728 Scott, S. K. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12),
729 2400–2406. <https://doi.org/10.1093/brain/123.12.2400>

730 Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception.
731 *Trends in Neurosciences*, 26(2), 100–107. [https://doi.org/10.1016/S0166-2236\(02\)00037-1](https://doi.org/10.1016/S0166-2236(02)00037-1)

732 Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., Huang, M.-X., Ding, N., & Gao, J.-H. (2018). The
733 Cortical Maps of Hierarchical Linguistic Structures during Speech Perception. *Cerebral Cortex*, 1–9.
734 <https://doi.org/10.1093/cercor/bhy191>

735 Siddiqi, S. H., Kording, K. P., Parvizi, J., & Fox, M. D. (2022). Causal mapping of human brain function. *Nature*
736 *Reviews Neuroscience*, 23(6), 361–375. <https://doi.org/10.1038/s41583-022-00583-8>

737 Silva, A. B., Liu, J. R., Zhao, L., Levy, D. F., Scott, T. L., & Chang, E. F. (2022). A Neurosurgical Functional
738 Dissection of the Middle Precentral Gyrus during Speech Production. *The Journal of Neuroscience*,
739 42(45), 8416–8426. <https://doi.org/10.1523/JNEUROSCI.1614-22.2022>

740 Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception.
741 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130297.
742 <https://doi.org/10.1098/rstb.2013.0297>

743 Sperber, C. (2020). Rethinking causality and data complexity in brain lesion-behaviour inference and its
744 implications for lesion-behaviour modelling. *Cortex*, 126, 49–62.
745 <https://doi.org/10.1016/j.cortex.2020.01.004>

746 Stephen, E. P., Li, Y., Metzger, S., Oganian, Y., & Chang, E. F. (2023). Latent neural dynamics encode temporal
747 context in speech. *Hearing Research*, 437, 108838. <https://doi.org/10.1016/j.heares.2023.108838>

748 Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent
749 responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, 150(1–2), 225–
750 244. [https://doi.org/10.1016/S0378-5955\(00\)00203-3](https://doi.org/10.1016/S0378-5955(00)00203-3)

751 Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-
752 invasive brain recordings. *Nature Neuroscience*. <https://doi.org/10.1038/s41593-023-01304-9>

753 Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2004). A Point Process Framework
754 for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects.
755 *Journal of Neurophysiology*, 93(2), 1074–1089. <https://doi.org/10.1152/jn.00697.2004>

756 Ufer, C., & Blank, H. (2023). Multivariate analysis of brain activity patterns as a tool to understand predictive
757 processes in speech perception. *Language, Cognition and Neuroscience*, 1–17.
758 <https://doi.org/10.1080/23273798.2023.2166679>

759 Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion Studies in Contemporary
760 Neuroscience. *Trends in Cognitive Sciences*, 23(8), 653–671. <https://doi.org/10.1016/j.tics.2019.05.009>

761 Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive*
762 *Science*, 25(1), 111–172. https://doi.org/10.1207/s15516709cog2501_5

763 Van Wassenhove, V. (2009). Minding time in an amodal representational space. *Philosophical Transactions of*
764 *the Royal Society B: Biological Sciences*, 364, 1815–1830. <https://doi.org/10.1098/rstb.2009.0023>

765 Venezia, J. H., Thurman, S. M., Richards, V. M., & Hickok, G. (2019). Hierarchy of speech-driven
766 spectrotemporal receptive fields in human auditory cortex. *NeuroImage*, 186, 647–666.
767 <https://doi.org/10.1016/j.neuroimage.2018.11.049>

768 Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation Through Neural Population Dynamics.
769 *Annual Review of Neuroscience*, 43(1), 249–275. <https://doi.org/10.1146/annurev-neuro-092619-094115>

770 Walker, K. M. M., Bizley, J. K., King, A. J., & Schnupp, J. W. H. (2011). Multiplexed and robust representations of
771 sound features in auditory cortex. *The Journal of Neuroscience: The Official Journal of the Society for*
772 *Neuroscience*, 31(41), 14565–14576. <https://doi.org/10.1523/JNEUROSCI.2074-11.2011>

773 Weichwald, S., & Jonas, P. (2021). Causality in Cognitive Neuroscience: Concepts, Challenges, and
774 Distributional Robustness. *Journal of Cognitive Neuroscience*, 33(2), 226–247.
775 https://doi.org/10.1162/jocn_a_01623

776 Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., & Grosse-Wentrup, M. (2015). Causal
777 interpretation rules for encoding and decoding models in neuroimaging. *NeuroImage*, 110, 48–59.
778 <https://doi.org/10.1016/j.neuroimage.2015.01.036>

779 Wernicke, C. (1874). *Der aphasische Symptomenkomplex. Eine psychologische Studie auf anatomischer Basis.*
780 Max Cohn & Weigert.

781 Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance
782 brain-to-text communication via handwriting. *Nature*, 593(7858), 249–254.
783 <https://doi.org/10.1038/s41586-021-03506-2>

784 Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., Kamdar, F., Glasser, M. F.,
785 Hochberg, L. R., Druckmann, S., Shenoy, K. V., & Henderson, J. M. (2023). A high-performance speech
786 neuroprosthesis. *Nature*, 620(7976), 1031–1036. <https://doi.org/10.1038/s41586-023-06377-x>

787 Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The Encoding of Speech Sounds in the Superior Temporal
788 Gyrus. *Neuron*, 102(6), 1096–1110. <https://doi.org/10.1016/j.neuron.2019.04.023>

789 Yourganov, G., Fridriksson, J., Rorden, C., Gleichgerrcht, E., & Bonilha, L. (2016). Multivariate Connectome-
790 Based Symptom Mapping in Post-Stroke Patients: Networks Supporting Language and Speech. *The*
791 *Journal of Neuroscience*, 36(25), 6668–6679. <https://doi.org/10.1523/JNEUROSCI.4396-15.2016>

792 Zhang, Q., Hu, X., Hong, B., & Zhang, B. (2019). A hierarchical sparse coding model predicts acoustic feature
793 encoding in both auditory midbrain and cortex. *PLoS Computational Biology*, 15(2), 1–23.
794 <https://doi.org/10.1371/journal.pcbi.1006766>

795 Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate lesion-symptom
796 mapping using support vector regression: Multivariate Lesion Symptom Mapping. *Human Brain*
797 *Mapping*, 35(12), 5861–5876. <https://doi.org/10.1002/hbm.22590>

798